

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> <b>C12Q 1/68, C12P 19/34, G05B 15/00, C07H 21/02, 21/04</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/58722</b> <b>(43) International Publication Date:</b> 18 November 1999 (18.11.99)
<b>(21) International Application Number:</b> PCT/US99/10510 <b>(22) International Filing Date:</b> 12 May 1999 (12.05.99)		<b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
<b>(30) Priority Data:</b> 60/085,092 12 May 1998 (12.05.98) US 09/076,447 12 May 1998 (12.05.98) US		<b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(71) Applicant</b> (for all designated States except US): ISIS PHARMACEUTICALS, INC. [US/US]; 2292 Faraday Avenue, Carlsbad, CA 92008 (US).			
<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants</b> (for US only): GRIFFEY, Richard [US/US]; 360 Barsby Street, Vista, CA 92084 (US). MOHAN, Venkatraman [IN/US]; 7042 Via Cabana, Carlsbad, CA 92009 (US).			
<b>(74) Agents:</b> CALDWELL, John, W. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).			
<b>(54) Title:</b> CHARACTERIZATION OF INTERACTIONS BETWEEN MOLECULAR INTERACTION SITES OF RNA AND LIGANDS THEREFOR			
<b>(57) Abstract</b> <p>The present invention is directed to methods of identifying compounds which bind to a molecular interaction site of a nucleic acid comprising providing a numerical representation of the three-dimensional structure of the molecular interaction site, providing a compound data set comprising numerical representations of the three-dimensional structures of a plurality of organic compounds, and comparing the numerical representation of the molecular interaction site with members of the compound data set to generate a hierarchy of organic compounds ranked in accordance with the ability of the organic compounds to form physical interactions with the molecular interaction site. Data sets comprising the numerical representations of the three-dimensional structures of molecular interaction sites, and comprising the numerical representations of the three-dimensional structure of a plurality of organic compounds are also described.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon			PT	Portugal		
CN	China	KR	Republic of Korea	RO	Romania		
CU	Cuba	KZ	Kazakhstan	RU	Russian Federation		
CZ	Czech Republic	LC	Saint Lucia	SD	Sudan		
DE	Germany	LI	Liechtenstein	SE	Sweden		
DK	Denmark	LK	Sri Lanka	SG	Singapore		
EE	Estonia	LR	Liberia				

## **CHARACTERIZATION OF INTERACTIONS BETWEEN MOLECULAR INTERACTION SITES OF RNA AND LIGANDS THEREFOR**

### **CROSS REFERENCE TO RELATED APPLICATIONS**

The present application is a continuation-in-part of U.S. Serial No. 09/076,447 filed 5 May 12, 1998, which claims priority to provisional U.S. Serial No. 60/085,092 filed May 12, 1998, each of which is incorporated herein by reference in its entirety.

### **FIELD OF THE INVENTION**

The present invention is directed to methods of identifying compounds which bind to molecular interaction sites of nucleic acids, especially RNA. The present invention 10 is also directed to the numerical representations of the three dimensional structures of molecular interaction sites and the compounds which interact with those sites.

### **BACKGROUND OF THE INVENTION**

The selection of compounds for synthesis and screening is a critical step in any drug discovery process. This is particularly true for combinatorial chemistry-based discovery 15 strategies, where a very much larger number of compounds can be conceived than can be prepared in a reasonable time frame. Computational chemistry methods have been applied to find the "best" sets of compounds for screening. One strategy optimizes the chemical "diversity" in a library in order to increase the likelihood of finding a hit with biological activity in a screen against a macromolecular target of unknown structure.

- 2 -

Targeting nucleic acids has been recognized as a valid strategy for interference with biological pathways and the treatment of disease. In this regard, both deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) have been the target of numerous therapeutic strategies. A wide variety of "small" molecules, oligomers and oligonucleotides have been 5 shown to possess binding affinity for nucleic acids. The vast majority of experience in interfering with nucleic acid function has been via the specific binding of ligands to a particular base, base pair, and/or primary sequence of bases in the nucleic acid target. Some compounds have also demonstrated a composite specificity that arises from recognition and interactions with both the primary and secondary structural features of the nucleic acid, such 10 as preferential binding to A-T base pairs in the DNA minor groove, with little or no binding to corresponding RNA sequences.

Exploiting the knowledge of the three-dimensional structure of biological targets is a promising strategy from a drug design and discovery standpoint. This has been demonstrated by the design and development of numerous drugs and drug candidates targeted 15 to proteins involved in various pathophysiological pathways. While three dimensional structures of proteins have been widely determined by techniques such as X-ray crystallography, molecular modeling and NMR, nucleic acid targets have been difficult to study. The literature reveals few three dimensional structures of biologically active RNA, including a tRNA, said to have been determined via X-ray crystallography. Quigley, *et al.*, 20 *Nucleic Acids Res.*, 1975, 2, 2329; and Moras, *et al.*, *Nature (London)*, 1980, 288, 669. The difficulties associated with proper crystallization and study of nucleic acids by X-ray methods along with the increasing number of biologically important small RNAs have increased the need for new structure determination and drug discovery strategies for such targets.

Many approaches to predicting RNA structure have been discussed in the 25 scientific literature. Essentially, these involve sequencing and genomic analysis of nucleic acids, such as RNA, as a first step to establish the primary sequence structure and potential folded structures of the target. A second step entails definition of structural constraints such as base pairing and long range interactions among bases based on information derived from cross-linking, biochemical and genetic structure-function studies. This information, together 30 with modeling and simulation software, has allowed scientists to predict three dimensional models of RNA and DNA. While such models may not be as powerful as X-ray crystal

- 3 -

structures, they have been useful in ascertaining some structural features and structure-function relationships.

An understanding of the structural features of specific motifs in nucleic acids, especially hairpins, loops, helices and double helices, has been found to be useful in gaining 5 molecular insights. For example, a hairpin motif comprising a double helical stem and a single-stranded loop is believed to be one of the simplest yet most important structural element in nucleic acids. Such hairpin structures are proposed to be nucleation sites and serve as major building blocks for the folded three dimensional structure of RNAs. Shen, *et al.*, *FASEB J.*, 1995, 9, 1023. Hairpins are also involved in specific interactions with a variety of 10 proteins to regulate gene expression. Feng, *et al.*, *Nature*, 1988, 334, 165, Witherell, *et al.*, *Prog. Nucleic Acids Res. Mol. Biol.*, 1991, 40, 185, and Phillippe, *et al.*, *J. Mol. Biol.*, 1990, 211, 415. Nucleic acid hairpin structures have therefore been widely studied by NMR, molecular modeling techniques such as constrained molecular dynamics and distance 15 geometry (Cheong, *et al.*, *Nature*, 1990, 346, 680 and Cain, *et al.*, *Nuc. Acids Res.*, 1995, 23, 2153), X-ray crystallography (Valegard, *et al.*, *Nature*, 1994, 371, 623 and Chatopadhyaya, *et al.*, *Nature*, 1988, 334, 175), and theoretical methods (Tung, *Biophysical J.*, 1997, 72, 876, Erie, *et al.*, *Biopolymers*, 1993, 33, 75, and Raghunathan, *et al.*, *Biochemistry*, 1991, 30, 782.

The determination of potential three dimensional structures of nucleic acids and their attendant structural motifs affords insights into areas such as the study of catalysis 20 by RNA, RNA-RNA interactions, RNA-nucleic acid interactions, RNA-protein interactions, and the recognition of small molecules by nucleic acids. Four general approaches to the generation of model three dimensional structures of RNA have been demonstrated in the literature. All of these employ sophisticated molecular modeling and computational algorithms for the simulation of folding and tertiary interactions within target nucleic acids, 25 such as RNA. Westhof and Altman (*Proc. Natl. Acad. Sci.*, 1994, 91, 5133) have described the generation of a three-dimensional working model of M1 RNA, the catalytic RNA subunit of RNase P from *E. coli* via an interactive computer modeling protocol. Leveraging the significant body of work in the area of cryo-electron microscopy (cryo-EM) and biochemical studies on ribosomal RNAs, Mueller and Brimacombe (*J. Mol. Biol.*, 1997, 271, 524) have 30 constructed a three dimensional model of *E. coli* 16S Ribosomal RNA. A method to model nucleic acid hairpin motifs has been developed based on a set of reduced coordinates for

- 4 -

describing nucleic acid structures and a sampling algorithm that equilibrates structures using Monte Carlo (MC) simulations (Tung, *Biophysical J.*, 1997, 72, 876, incorporated herein by reference in its entirety). MC-SYM is yet another approach to predicting the three dimensional structure of RNAs using a constraint-satisfaction method. Major, *et al.*, *Proc. Natl. Acad. Sci.*, 1993, 90, 9408. The MC-SYM program is an algorithm based on constraint satisfaction that searches conformational space for all models that satisfy query input constraints, and is described in, for example, Cedergren, *et al.*, *RNA Structure And Function*, 1998, Cold Spring Harbor Lab. Press, p.37-75. Three dimensional structures of RNA are produced by this method by the stepwise addition of a nucleotide having one or several 10 different conformations to a growing oligonucleotide model.

Westhof and Altman (*Proc. Natl. Acad. Sci.*, 1994, 91, 5133) have described the generation of a three-dimensional working model of M1 RNA, the catalytic RNA subunit of RNase P from *E. coli* via an interactive computer modeling protocol. This modeling protocol incorporated data from chemical and enzymatic protection experiments, phylogenetic 15 analysis, studies of the activities of mutants and the kinetics of reactions catalyzed by the binding of substrate to M1 RNA. Modeling was performed for the most part as described in the literature. Westhof, *et al.*, in "Theoretical Biochemistry and Molecular Biophysics," Beveridge and Lavery (eds.), Adenine, NY, 1990, 399. In general, starting with the primary sequence of M1 RNA, the stem-loop structures and other elements of secondary structure 20 were created. Subsequent assembly of these elements into a three dimensional structure using a computer graphics station and FRODO (Jones, *J. Appl. Crystallogr.*, 1978, 11, 268) followed by refinement using NUCLIN-NUCLSQ afforded a RNA model that had correct geometries, the absence of bad contacts, and appropriate stereochemistry. The model so generated was found to be consistent with a large body of empirical data on M1 RNA and 25 opens the door for hypotheses about the mechanism of action of RNase P. However, the models generated by this method are less well resolved than the structures determined via X-ray crystallography.

Mueller and Brimacombe (*J. Mol. Biol.*, 1997, 271, 524) have constructed a three dimensional model of *E. coli* 16S ribosomal RNA using a modeling program called 30 ERNA-3D. This program generates three dimensional structures such as A-form RNA helices and single-strand regions via the dynamic docking of single strands to fit electron density

- 5 -

obtained from low resolution diffraction data. After helical elements have been defined and positioned in the model, the configurations of the single strand regions is adjusted, so as to satisfy any known biochemical constraints such as RNA-protein cross-linking and footprinting data.

5        A method to model nucleic acid hairpin motifs has been developed based on a set of reduced coordinates for describing nucleic acid structures and a sampling algorithm that equilibrates structures using Monte Carlo (MC) simulations. Tung, *Biophysical J.*, 1997, 72, 876, incorporated herein by reference. The stem region of a nucleic acid can be adequately modeled by using a canonical duplex formation. Using a set of reduced 10 coordinates, an algorithm that is capable of generating structures of single stranded loops with a pair of fixed ends was created. This allows efficient structural sampling of the loop in conformational space. Combining this algorithm with a modified Metropolis Monte Carlo algorithm afforded a structure simulation package that simplifies the study of nucleic acid hairpin structures by computational means.

15        Knowledge and mastery of the foregoing techniques is assumed to be part of the ordinary skill in the art. There has been a long-felt need in the art to provide methods for improved determination of the three-dimensional structure of important regulatory and other elements in nucleic acids, especially RNA. It is also been greatly desired to achieve improved knowledge about the nature of interactions between ligands or potential ligands and nucleic 20 acids, especially RNA. The present invention is directed towards satisfaction of these objectives.

Accordingly, it is an objective of the present invention to provide improved characterization of interactions between RNA and other nucleic acids and ligands or potential ligands therefor.

25        A further object of the invention is to compare molecular interaction sites of RNA with compounds proposed for interaction therewith.

In accordance with preferred embodiments of the present invention, the comparison of molecular interaction sites of RNA with compounds is achieved through comparison of numerical representations of the three-dimensional structure of the molecular 30 interaction site with the three dimensional structure of the ligands in a fashion such that such interactions can be compared as to quality.

- 6 -

Another object of the present invention is the preparation of hierarchies of ligands ranked or ordered in accordance with their ability to interact with molecular interaction sites of RNA and other nucleic acid targets.

Yet another object of the present invention is the establishment of databases 5 of the numerical representations of three-dimensional structures of molecular interaction sites of nucleic acids and three-dimensional structures of libraries of ligands. Such databases libraries provide powerful tools for the elucidation of structure and interactions of molecular interaction sites with potential ligands and predictions thereof.

Other objectives will become apparent to persons of ordinary skill in the art 10 upon review of the present specification and appended claims.

## SUMMARY OF THE INVENTION

The present invention is directed to methods of identifying compounds which bind to a molecular interaction site of a nucleic acid comprising providing a numerical 15 representation of the three-dimensional structure of the molecular interaction site and providing a compound data set comprising numerical representations of the three dimensional structures of a plurality of organic compounds. The numerical representation of the molecular interaction site is then compared with members of the compound data set to generate a hierarchy of organic compounds ranked in accordance with the ability of the organic 20 compounds to form physical interactions with the molecular interaction site.

The present invention is also directed to data sets comprising the numerical representations of the three dimensional structures of molecular interaction sites and to the numerical representations of the three dimensional structure of a plurality of organic compounds.

25 The present invention is directed to methods of identifying compounds which bind to a molecular interaction site of nucleic acids. They comprise providing a numerical representation of the three dimensional structure of the molecular interaction site, providing a compound data set comprising numerical representations of the three dimensional structures of a plurality of organic compounds, comparing the numerical representation of the molecular 30 interaction site with members of the compound data set to generate a hierarchy of organic

- 7 -

compounds which is ranked in accordance with the ability of the organic compounds to form physical interactions with the molecular interaction site.

While there are a number of ways to identify molecular interaction sites, identify compounds likely to interact with molecular interaction sites of RNA and other 5 biological molecules, synthesize such compounds and analyze their binding, preferred methodologies are described in U.S. Serial Numbers 09/076,440, 09/076,405, 09/076,447, 09/076,206, 09/076,214, and 09/076,404, each of which was filed on May 12, 1998 and each assigned to the assignee of this invention. All of the foregoing applications are incorporated by reference herein in their entirety.

## 10 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows exemplary compounds which were docked to TAR with subsequent evaluation of the solvation/desolvation energy.

Figure 2 shows the target RNA for 4.5S-P48.

Figure 3A shows a representative demonstration of cap-dependent translation 15 of three DNA plasmids with a wheat germ lysate system: a) a luciferase gene with a 9 base leader sequence before the AUG start codon; b) translation of a construct with the TAR RNA structure adjacent to the cap; c) translation of a construct with the TAR RNA structure separated from the cap by a 9 base leader sequence. Solid bars: no added m<sup>7</sup>G. Hatched bars: added m<sup>7</sup>G.

20 Figure 3B shows an exemplary inhibition of translation of an mRNA construct containing the TAR RNA structure by a 39 amino acid *tat* peptide: a) translation of a luciferase mRNA with a 9 base leader sequence with and without 10  $\mu$ M added *tat* peptide; b) translation of luciferase mRNA containing the TAR RNA structure adjacent to the cap; c) translation of the luciferase/TAR RNA construct with a 9 base leader in the presence/absence 25 of 10  $\mu$ M *tat* peptide.

Figure 4 shows an exemplary dose-dependent inhibition of translation of a 30 luciferase mRNA construct containing a TAR RNA structure in the 5'-UTR by ACD 00001199 (DecpBlue-3). Solid line: inhibition of translation of the control luc+9 plasmid. Dashed line: inhibition of expression of the luc+9 mRNA containing the TAR RNA structure of the 5'-UTR.

- 8 -

Figure 5 shows a representative lowest energy structure of paromomycin (dark grey) bound to bacterial 16S ribosomal A site (not shown) identified using the QXP method for the lowest energy conformers. The target RNA was held rigid whereas the paromomycin was treated as fully flexible. The structure obtained using NMR is shown in light grey.

5 Figure 6 shows a representative correlation between the observed rms deviation and QXP energy scores obtained for the bacterial 16S ribosomal A site bound to paromomycin. I1-I5 represent separate runs.

## DETAILED DESCRIPTION OF THE INVENTION

A molecular interaction site is a region of a nucleic acid which has secondary 10 structure. Preferably, the molecular interaction site is conserved between a plurality of different taxonomic species. The nucleic acid can be either eukaryotic or prokaryotic. The nucleic acid is preferably mRNA, pre-mRNA, tRNA, rRNA, or snRNA. The RNA can be viral, fungal, parasitic, bacterial, or yeast. Preferably, the molecular interaction site is present in a region of an RNA which is highly conserved among a plurality of taxonomic species.

15 Molecular interaction site are described in further detail in U.S. Application Serial No. 09/076,440, filed May 12, 1998, which is assigned to the assignee of the present application, which is incorporated herein by reference in its entirety. In accordance with some preferred embodiments of this invention, it will be appreciated that the biomolecules having a molecular interaction site or sites, especially RNAs, may be derived from a number of sources. Thus,

20 such RNA targets can be identified by any means, rendered into three dimensional representations and employed for the identification of compounds which can interact with them to effect modulation of the RNA.

The three dimensional structure of a molecular interaction site, preferably of an RNA, can be manipulated as a numerical representation. Computer software that provides 25 one skilled in the art with the ability to design molecules based on the chemistry being performed and on available reaction building blocks is commercially available. Software packages from companies such as, for example, Tripos (St. Louis, MO), Molecular Simulations (San Diego, CA), MDL Information Systems (San Leandro, CA) and Chemical Design (NJ) provide means for computational generation of structures. These software 30 products also provide means for evaluating and comparing computationally generated

molecules and their structures. *In silico* collections of molecular interaction sites can be generated using the software from any of the above-mentioned vendors and others which are or may become available

A set of structural constraints for the molecular interaction site of the RNA can 5 be generated from biochemical analyses such as, for example, enzymatic mapping and chemical probes, and from genomics information such as, for example, covariance and sequence conservation. Information such as this can be used to pair bases in the stem or other region of a particular secondary structure. Additional structural hypotheses can be generated for noncanonical base pairing schemes in loop and bulge regions. A Monte Carlo search 10 procedure can sample the possible conformations of the RNA consistent with the program constraints and produce three dimensional structures.

Reports of the generation of three dimensional, *in silico* representations are available from the standpoint of library design, generation, and screening against protein targets. Likewise, some efforts in the area of generating RNA models have been reported in 15 the literature. However, there are no reports on the use of structure-based design approaches to query *in silico* representations of organic molecules, "small" molecules, oligonucleotides or other nucleic acids, with three dimensional, *in silico*, representations of RNA structures. The present invention preferably employs computer software that allows the construction of three dimensional models of RNA structure, the construction of three dimensional, *in silico* 20 representations of a plurality of organic compounds, "small" molecules, polymeric compounds, oligonucleotides and other nucleic acids, screening of such *in silico* representations against RNA molecular interaction sites *in silico*, scoring and identifying the best potential binders from the plurality of compounds, and finally, synthesizing such compounds in a combinatorial fashion and testing them experimentally to identify new ligands 25 for such targets.

In preferred embodiments of the invention, an automated computational search algorithm, such as those described above, is used to predict all of the allowed three dimensional molecular interaction site structures, preferably from RNA, which are consistent with the biochemical and genomic constraints specified by the user. Based, for example on 30 their root-mean-squared deviation values, these structures are clustered into different families.

- 10 -

A representative member or members of each family can be subjected to further structural refinement via molecular dynamics with explicit solvent and cations.

Structural enumeration and representation by these software programs is typically done by drawing molecular scaffolds and substituents in two dimensions. Once 5 drawn and stored in the computer, these molecules may be rendered into three dimensional structures using algorithms present within the commercially available software. Preferably, MC-SYM is used to create three dimensional representations of the molecular interaction site. The rendering of two dimensional structures of molecular interaction sites into three dimensional models typically generates a low energy conformation or a collection of low 10 energy conformers of each molecule. The end result of these commercially available programs is the conversion of a nucleic acid sequence containing a molecular interaction site into families of similar numerical representations of the three dimensional structures of the molecular interaction site. These numerical representations form an ensemble data set.

The three dimensional structures of a plurality of compounds, preferably 15 "small" organic compounds, can be designated as a compound data set comprising numerical representations of the three dimensional structures of the compounds. "Small" molecules in this context refers to non-oligomeric organic compounds. Two dimensional structures of compounds can be converted to three dimensional structures, as described above for the molecular interaction sites, and used for querying against three dimensional structures of the 20 molecular interaction sites. The two dimensional structures of compounds can be generated rapidly using structure rendering algorithms commercially available. The three dimensional representation of the compounds which are polymeric in nature, such as oligonucleotides or other nucleic acids structures, may be generated using the literature methods described above. A three dimensional structure of "small" molecules or other compounds can be generated and 25 a low energy conformation can be obtained from a short molecular dynamics minimization. These three dimensional structures can be stored in a relational database. The compounds upon which three dimensional structures are constructed can be proprietary, commercially available, or virtual.

In some preferred embodiments of the invention, a compound data set 30 comprising numerical representations of the three dimensional structure of a plurality of organic compounds is provided by, for example, Converter (MSI, San Diego) from two

dimensional compound libraries generated by, for example, a computer program modified from commercial programs. Other suitable databases can be constructed by converting two dimensional structures of chemical compounds into three dimensional structures, as described above. The software is described in greater detail in U.S. Application Serial No. 09/076,405, 5 filed May 12, 1998, which is assigned to the assignee of the present application, and which is incorporated herein by reference in its entirety. The end result is the conversion of two dimensional structures of organic compounds into numerical representations of the three dimensional structures of a plurality of organic compounds. These numerical representations are presented as a compound data set.

10 After both the numerical representations of the three-dimensional structure of the molecular interaction sites and the compound data set comprising numerical representations of the three dimensional structures of a plurality of organic compounds are obtained, the numerical representations of the molecular interaction sites are compared with members of the compound data set to generate a hierarchy of the organic compounds. The 15 hierarchy is ranked in accordance with the ability of the organic compounds to form physical interactions with the molecular interaction site. Preferably, the comparing is carried out *seriatim* upon the members of the compound data set. In accordance with some embodiments, the comparison can be performed with a plurality of molecular interaction sites at the same time.

20 A variety of theoretical and computational methods are known by those skilled in the art to study and optimize the interactions of "small" molecules or organic compounds with biological targets such as nucleic acids. These structure-based drug design tools have been very useful in modeling the interactions of proteins with small molecule ligands and in optimizing these interactions. Typically this type of study has been performed when the 25 structure of the protein receptor was known by querying individual small molecules, one at a time, against this receptor. Usually these small molecules had either been co-crystallized with the receptor, were related to other molecules that had been co-crystallized or were molecules for which some body of knowledge existed concerning their interactions with the receptor. A significant advance in this area was the development of a software program called 30 DOCK that allows structure-based database searches to find and identify molecules that are expected to bind to a receptor of interest. Kuntz, *et al.*, *Acc. Chem. Res.*, **1994**, 27, 117, and

- 12 -

Gschwend and Kuntz, *J. Compt.-Aided Mol. Des.*, 1996, 10, 123. DOCK 4.0 is commercially available from the Regents of the University of California. Equivalent programs are also comprehended in the present invention. DOCK allows the screening of a large collection of molecules whose three dimensional structures have been generated *in silico*, *i.e.*, in computer readable format, but for which no prior knowledge of interactions with the ligands is available. DOCK, therefore, is a significant tool to the process of discovering new ligands to a molecule of interest and is presently preferred for use herein.

The DOCK program has been widely applied to protein targets and the identification of ligands that bind to them. Typically, new classes of molecules that bind to known targets have been identified, and later verified by *in vitro* experiments. The DOCK software program consists of several modules, including SPHGEN (Kuntz, *et al.*, *J. Mol. Biol.*, 1982, 161, 269) and CHEMGRID (Meng, *et al.*, *J. Comput. Chem.*, 1992, 13, 505). SPHGEN generates clusters of overlapping spheres that describe the solvent-accessible surface of the binding pocket within the target receptor. Each cluster represents a possible binding site for small molecules. CHEMGRID precalculates and stores in a grid file the information necessary for force field scoring of the interactions between binding molecule and target. The scoring function approximates molecular mechanics interaction energies and consists of van der Waals and electrostatic components. DOCK uses the selected cluster of spheres to orient ligands molecules in the targeted site on the receptor. Each molecule within a previously generated three dimensional database is tested in thousands of orientations within the site, and each orientation is evaluated by the scoring function. Only that orientation with the best score for each compound so screened is stored in the output file. Finally, all compounds of the database are ranked in a hierarchy, *e.g.*, ordered by scores, and a collection of the best candidates may then be screened experimentally.

Using DOCK, numerous ligands have been identified for a variety of protein targets. Recent efforts in this area have resulted in reports of the use of DOCK to identify and design small molecule ligands that exhibit binding specificity for nucleic acids such as RNA double helices. While RNA plays a significant role in many diseases such as AIDS, viral and bacterial infections, few studies have been made on small molecules capable of specific RNA binding. Compounds possessing specificity for the RNA double helix, based on the unique geometry of its deep major groove, were identified using the DOCK methodology. Chen, *et*

- 13 -

al., *Biochemistry*, 1997, 36, 11402 and Kuntz, et al., *Acc. Chem. Res.*, 1994, 27, 117. Recently, the application of DOCK to the problem of ligand recognition in DNA quadruplexes has been reported. Chen, et al., *Proc. Natl. Acad. Sci.*, 1996, 93, 2635.

Programs such as DOCK typically assume knowledge of the conformation of the bound ligand and use a rigid conformation for a given ligand in molecular docking studies to arrive at structures of ligand-receptor complexes (which is a prerequisite for computing binding energies). Most ligands, however, possess a number of rotatable bonds, thus increasing the complexity of the calculations. Docking of flexible ligands would be desirable, but requires one to search an enormous amount of conformational space. For example, the study of an aminoglycoside antibiotic (paromomycin) bound to 16S A-site RNA target, would constitute a search space of  $\sim 10^{30}$  possible solutions.

QXP is a method that permits flexible ligand docking calculations (McMartin, C. and Bohacek, R.S., *J. Comput.-Aided Mole. Design*, 1997, 11, 333). In this method, full conformational searches on flexible ligands are carried out. QXP search algorithms employ the Monte Carlo perturbation technique with energy minimization in Cartesian space. An additional fast search step is introduced between the initial perturbation and energy minimization. This method is also presently preferred for use herein.

Preferably, individual compounds to be used in these methods are designated as *mol* files, for example, and combined into a collection of *in silico* representations using appropriate computer software, such as the software described in greater detail in U.S. Application Serial No. 09/076,405, filed May 12, 1998, which is assigned to the assignee of the present application, and which is incorporated herein by reference in its entirety. These two dimensional *mol* files are exported and converted into three dimensional structures using commercial software such as Converter (Molecular Simulations Inc., San Diego) or equivalent software, as described above. Atom types suitable for use with a docking program such as DOCK or QXP are assigned to all atoms in the three dimensional *mol* file using software such as, for example, Babel, or with other equivalent software.

A low-energy conformation of each molecule is generated with software such as Discover (MSI, San Diego). An orientation search is performed by bringing each compound of the plurality of compounds into proximity with the molecular interaction site in many orientations using DOCK or QXP. A contact score is determined for each

- 14 -

orientation, and the optimum orientation of the compound is subsequently used. Alternatively, the conformation of the compound can be determined from a template conformation of the scaffold determined previously.

The interaction of a plurality of compounds and molecular interaction sites is  
5 examined by comparing the numerical representations of the molecular interaction sites with  
members of the compound data set. Preferably, a plurality of compounds such as those  
generated by computer programs or otherwise, is compared to the molecular interaction site  
and allowed to undergo random "motions" among the dihedral bonds of the compounds.  
Preferably about 20,000 to 100,000 compounds are compared to at least one molecular  
10 interaction site. Typically, 20,000 compounds are compared to about five molecular  
interaction sites and scored. Individual conformations of the three dimensional structures are  
placed at the target site in many orientations. Moreover, during execution of the DOCK  
program, the compounds and molecular interaction sites are allowed to be "flexible" such that  
the optimum hydrogen bonding, electrostatic, and van der Waals contacts can be realized.  
15 The energy of the interaction is calculated and stored for 10-15 possible orientations of the  
compounds and molecular interaction sites. QXP methodology allows true flexibility in both  
the ligand and target and is presently preferred.

The relative weights of each energy contribution are updated constantly to  
insure that the calculated binding scores for all compounds reflect the experimental binding  
20 data. The binding energy for each orientation is scored on the basis of hydrogen bonding, van  
der Waals contacts, electrostatics, solvation/desolvation, and the quality of the fit. The  
lowest-energy van der Waals, dipolar, and hydrogen bonding interactions between the  
compound and the molecular interaction site are determined, and summed. In preferred  
embodiments, these parameters can be adjusted according to the results obtained empirically.  
25 The binding energies for each molecule against the target are output to a relational database.  
The relational database contains a hierarchy of the compounds ranked in accordance with the  
ability of the compounds to form physical interactions with the molecular interaction site. The  
higher ranked compounds are better able to form physical interactions with the molecular  
interaction site.

30 In a preferred embodiment, the highest ranking, *i.e.*, the best fitting  
compounds, are selected for synthesis. In preferred embodiments of the invention, those

- 15 -

compounds which are likely to have desired binding characteristics based on binding data are selected for synthesis. Preferably the highest ranking 5% are selected for synthesis. More preferably, the highest ranking 10% are selected for synthesis. Even more preferably, the highest ranking 20% are selected for synthesis. The synthesis of the selected compounds can 5 be automated using a parallel array synthesizer or prepared using solution-phase or other solid-phase methods and instruments. In addition, the interaction of the highly ranked compounds with the nucleic acid containing the molecular interaction site is assessed as described below.

The interaction of the highly ranked organic compounds with the nucleic acid 10 containing the molecular interaction site can be assessed by numerous methods known to those skilled in the art. For example, the highest ranking compounds can be tested for activity in high-throughput (HTS) functional and cellular screens. HTS assays for each target RNA can be determined by scintillation proximity, precipitation, luminescence-based formats, filtration based assays, colorometric assays, and the like. Lead compounds can then be scaled 15 up and tested in animal models for activity and toxicity. The assessment preferably comprises mass spectrometry of a mixture of the nucleic acid and at least one of the compounds or a functional bioassay.

Certain preferred evaluation techniques employing mass spectroscopy are disclosed in U.S. Patent Application Serial. No. 09/076,206 filed May 12, 1998, which is 20 assigned to the assignee of the present application. The foregoing patent application is incorporated herein by reference in its entirety as exemplary of certain useful and preferred mass spectrometric techniques for use herewith. It is to be specifically understood, however, that it is not essential that these particular mass spectrometric techniques be employed in order to perform the present invention. Rather, any evaluative technique may be undertaken so long 25 as the objectives of the present invention are maintained.

In some embodiments of the invention, the highest ranking 20% of compounds from the hierarchy generated using the DOCK program or QXP are used to generate a further data set of three dimensional representations of organic compounds comprising compounds which are chemically related to the compounds ranking high in the hierarchy. Although the 30 best fitting compounds are likely to be in the highest ranking 1%, additional compounds, up to about 20%, are selected for a second comparison so as to provide diversity (ring size, chain

length, functional groups). This process insures that small errors in the molecular interaction sites are not propagated into the compound identification process. The resulting structure/score data from the highest ranking 20%, for example, is studied mathematically (clustered) to find trends or features within the compounds which enhance binding. The 5 compounds are clustered into different groups. Chemical synthesis and screening of the compounds, described above, allows the computed DOCK or QXP scores to be correlated with the actual binding data. After the compounds have been prepared and screened, the predicted binding energy and the observed Kd values are correlated for each compound.

The results are used to develop a predictive scoring scheme, which weighs 10 various factors (steric, electrostatic) appropriately. The above strategy allows rapid evaluation of a number of scaffolds with varying sizes and shapes of different functional groups for the high ranked compounds. In this manner, a further data set of representations of organic compounds comprising compounds which are chemically related to the organic compounds which rank high in the hierarchy can be compared to the numerical representations of the 15 molecular interaction site to determine a further hierarchy ranked in accordance with the ability of the organic compounds to form physical interactions with the molecular interaction site. In this manner, the further data set of representations of the three dimensional structures of compound which are related to the compounds ranked high in the hierarchy are produced and have, in effect, been optimized by correlating actual binding with virtual binding. The 20 entire cycle can be iterated as desired until the desired number of those compounds highest in the hierarchy are produced.

Compounds which have been determined to have affinity and specificity for a target biomolecule, especially a target RNA or which otherwise have been shown to be able to bind to the target RNA to effect modulation thereof, can, in accordance with preferred 25 embodiments of this invention, be tagged or labeled in a detectable fashion. Such labeling may include all of the labeling forms known to persons of skill in the art such as fluorophore, radiolabel, enzymatic label and many other forms. Such labeling or tagging facilitates detection of molecular interaction sites and permits facile mapping of chromosomes and other useful processes.

**EXAMPLES****Example 1: Functional Screening**

The compounds are screened for binding affinity using MASS or conventional high-throughput functional screens. The best scoring compounds from docking a 256-member library against the 16S A-site ribosomal RNA structure are shown in the table below. The DOCK scores ranged from -308.8 to -144.2 as listed in Table 1. The MASS assay was performed with the 27-mer model RNA sequence of the 16S A-site whose NMR structure has been determined. The transcription/translation assay was based on expression of a luciferase plasmid.

10      **Table 1. DOCK scores correlated with mass spectrometry and biological assay**

<b>Compound</b>	<b>DOCK score</b>	<b>MASS K<sub>D</sub></b>	<b>Activity<sup>1</sup></b>
Paromomycin	-308.8	0.5 $\mu$ M	0.3 $\mu$ M
170046	-303.4	>50	>100
169999	-299.0	>50	>100
15      169963	-293.9	>50	>100
170070	-290.2	>50	>100
169970	-288.9	1.5	2.5
169961	-288.5	5.0	10
20      170003	-287.8	>50	>100
169995	-286.4	>50	>100
169993	-286.0	>50	>100
170072	-282.6	>50	>100
170078	-281.6	5.0	10
25      169985	-280.1	4.0	10
169998	-278.0	>50	>100

<sup>1</sup>Inhibition of protein synthesis in transcription/translation assay for luciferase reporter.

Paromomycin is an aminoglycoside antibiotic known to bind to the A-site RNA structure. The NMR structure was determined with paromomycin bound at the A-site.

Paromomycin had the best DOCK contact score, along with high chemical and energy scores. The docking results for these compounds have been correlated with their binding affinity for a 16S RNA fragment using MASS mass spectrometry, and their ability to inhibit protein synthesis in a transcription/translation assay. Four of the 12 compounds with the best DOCK scores had good affinity (<10  $\mu$ M) for the RNA in the MASS assay and inhibited translation of a luciferase plasmid at <10  $\mu$ M. In addition, all 9 of the "good" binders in the MASS assay scored in the top 30% in the DOCK calculation.

Ibis compound 169970 had the best energy score of any compound, but had a poor contact score. This result suggests that the biological activity may be increased further 10 by modifying the structure to increase the number of close contacts with the 16S A-site RNA.

#### **Example 2: Target Site of TAR**

The NMR solution structure of TAR RNA (Varani, *et al.*, *J. Mol. Biol.*, 1995, 253, 313) has been used in the study of virtual screening for HIV-1 TAR RNA ligands. The compounds present in the Available Chemicals Database (ACD) have been partitioned into 15 a number of subsets according to their formal charges (neutral, +1, +2, etc) and DOCKed to the TAR structure. Five aminoglycoside antibiotics were among the 20 compounds with the best binding energies.

In addition, a number of compounds were docked to TAR with subsequent evaluation of the solvation/desolvation energy. An exemplary result is illustrated in Figure 20 1 which shows that ACD 00001199 and ACD 00192509 show relatively low energies of solvation/desolvation as well as low IC<sub>50</sub> values.

#### **Example 3: L11/Thiostrepton - An Example Of A High Throughput RNA/Protein Assay**

RNA molecules play numerous roles in cellular functions that range from structural to enzymatic in nature. These RNA molecules may work as single large molecules, 25 in complexes with one or more proteins, or in partnership with one or more RNA molecules. Some of these complexes, such as those found in the ribosome, have been virtually intractable as high throughput screening targets due to their immense size and complexity. The ribosome presents a particularly rich source of RNA structures and functions that would appear, at first glance, to be highly effective drug targets. A large number of natural antibiotics exist that are

directed against ribosomal targets indicating the general success of this strategy. These include the aminoglycosides, kirromycin, neomycin, paromomycin, thiostrepton, and many others. Thiostrepton, a cyclic peptide based antibiotic, inhibits several reactions at the ribosomal GTPase center of the 50S ribosomal subunit. Evidence exists that thiostrepton acts 5 by binding to the 23S rRNA component of the 50S subunit at the same site as the large ribosomal protein L11. The binding of L11 to the 23S rRNA causes a large conformation shift in the proteins tertiary structure. The binding of thiostrepton to the rRNA appears to cause an increase in the strength of the L11/23S rRNA interactions and prevents a conformational transition event in the L11 protein thereby stalling translation. Unfortunately, 10 thiostrepton has very poor solubility, relatively high toxicity, and is not generally useful as an antibiotic. The discovery of new, novel, antibiotics directed against these types of targets would be of great value.

The design of high throughput assays to discover new antibiotics directed against ribosomal targets has been difficult, in part, due to the large structures involved and 15 the low binding affinity of the RNA/protein interactions. Recently, a tremendous amount of data has been generated concerning RNA structures in the ribosome. This data has allowed the elucidation a number of structures and enabled the prediction of many others. Further, the use of the SPA assay format, as described below, allows for assays to be run without washing or other steps that lower the concentrations of binding components. This allows one to 20 examine binding interactions with very low ( $> 1 \mu\text{M}$ ) Kd's.

The mode of action of thiostrepton appears to be to stabilize a region of the 23S rRNA and by doing so prevent a structural transition in the L11 protein. Among the many assays that look at RNA/protein interactions, a SPA assay has been designed to look for small molecules that could be effective as thiostrepton 'like' agents. This assay uses a radiolabeled 25 small fragment of the 23S rRNA, a biotinylated 75 amino acid fragment of the L11 protein that contains the 23S rRNA binding domain and thiostrepton. The folding conditions of the secondary and tertiary structures of the 23S rRNA fragment have been examined as have the binding conditions of the L11 fragment to the 23S rRNA. The L11-thiostrepton assay has been optimized so that the 23S rRNA fragment is in an unfolded state prior to the addition of 30 compounds. Addition of the L11 fragment to this unfolded RNA results in no detectable binding interaction. The high throughput assay is run by mixing the 23S rRNA fragment,

- 20 -

under destabilizing conditions, with compounds of interest, incubating this mixture, and then adding the L11 fragment. Streptavidin-coated SPA beads are added for binding detection. Thiomestrepton is used as a positive control. Addition of thiostrepton to the RNA promotes the correct secondary and/or tertiary folding of the structure and allows the L11 fragment to bind 5 leading to the generation of a signal in the assay.

A tested paradigm has been developed for designing, developing and performing high and low throughput assays to look at RNA/protein function, structure, and binding in bacteria. The L11/thiomestrepton assay described above is but one of a number of RNA/protein interaction and functional assays that we have designed and developed for high 10 and low throughput screening. Others include functional assays to measure RnaseP, RnaseE, and EF-Tu activity. Assays to examine the function of the bacterial signal recognition particle and S30 assembly are also contemplated.

#### Example 4: P48-4.5S Interaction

The P48 protein-binding region of the 4.5S RNA present in the signal 15 recognition particle of bacteria has been selected as a target. The binding of P48 to 4.5S RNA is essential for bacteria to survive, and development of an inhibitor of this binding should generate a novel class of antimicrobial agent. Using compounds ( $\sim 2 \times 10^5$ ) from the Available Chemicals Directory (ACD), as well as from additional libraries, initial screening using DOCK (Meng, *et al.*, *J. Comp. Chem.*, 1992, 13, 505-524, incorporated herein by reference 20 in its entirety) (version 4.0) can be carried out. This should leave about 15-20% of the compounds in the database which have reasonably good shape complementarity in docking to the NMR structure of the 46mer, which is from the asymmetric bulged regions of *E. coli* 4.5S RNA. A pseudobrownian Monte Carlo search in torsion angle space is performed using the 25 program ICM (version 2.6), coupled with local minimization of each conformation, for automated flexible docking of that truncated set of potential ligands to the NMR structure and scoring for predicted affinity using an empirical free energy function.

Approximately 2000 of the best scoring compounds will be examined for experimental testing of their capability to inhibit the binding of P48 to 4.5S RNA. Inhibition of P48-4.5S RNA binding produced by the selected compounds will be measured using (his)<sub>6</sub>- 30 tagged P48 and <sup>33</sup>P-RNA in a high-throughput scintillation proximity assay system. The

structure-activity relationship among these 2000 compounds will serve as the basis for an expanded synthetic effort.

Docking of small molecules to the region of the asymmetric RNA bulges is expected to identify compounds with a high probability of selectively destabilizing the 4.5S-5 P48 interaction *in vitro*. The structure for the target RNA, shown in Figure 2, will be determined using NMR. Compounds (approaching  $2 \times 10^5$ ) from the Available Chemicals Directory (ACD) will be docked to the structure and scored for predicted affinity. The best molecules will be screened for their ability to disrupt the RNA-protein interaction. Quantitative structure-activity relationship (QSAR) studies will be performed on the most active compounds to identify critical features and interactions with the RNA. New compounds (~20,000) will be prepared through combinatorial addition and/or repositioning of hydrogen bonding, aromatic, and charged functional groups to enhance the activity and specificity of the compounds for the bacterial SRP relative to the human counterpart. In addition, a pseudobrownian Monte Carlo search in torsion angle space using the program ICM2.6 (Abagyan, *et al.*, *J. Comp. Chem.*, 1994, 15, 488-506, incorporated herein by reference in its entirety) will be performed, coupled with local minimization of each conformation, for automated flexible docking of the truncated database to the NMR structural models.

In order to rank the ligands after flexible docking is completed, a function to estimate their binding free energies is used. There are a number of empirical methods for estimation of the free energy of binding, but we intend to use the empirical free energy function we derived from the thermodynamic binding cycle (Filikov, *et al.*, *J. Comp.-Aided Molec. Design*, 1998, 12, 1-12, which is incorporated herein by reference in its entirety).

**Example 5: Inhibition of Translation of an mRNA Containing a Molecular Interaction Site by a "Small" Molecule Identified by Molecular Docking**

Translation of mRNAs in eukaryotic cells follows formation of an initiation complex at the 5'-cap ( $m^7Gppp$ ). A variety of initiation factors bind to the 5'-cap to form a pre-initiation complex before the 40S ribosomal subunit binds to the 5'-untranslated region upstream of the AUG start codon. Pain, *Eur. J. Biochem.*, 1996, 236, 747-771. It has been demonstrated that RNA secondary structures near the 5'-cap can affect the rates of translation

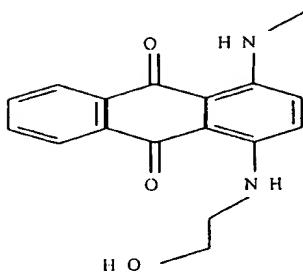
- 22 -

of mRNAs. Kozak, *J. Biol. Chemistry*, **1991**, 266, 19867-19870. These RNA structures can bind proteins and inhibit the level of translation. Standart, *et al.*, *Biochimie*, **1994**, 76, 867-879. The translational machinery has an ATP-dependant RNA helicase activity associated with the eIF-4a/eIF-4b complex, and under normal conditions, the RNA structures are opened 5 by the helicase and do not slow the rate of translation of the mRNA. The eIF-4a has a low, i.e.,  $\mu$ M, affinity for the pre-initiation complex.

It is believed that stabilization of mRNA structures near the 5'-cap also could be effected by specific "small" molecules, and that such binding would reduce the translational efficiency of the mRNA. To test this hypothesis, a plasmid was constructed 10 containing the luciferase message behind a 5'-UTR containing a 27-mer RNA construct of the HIV TAR stem-loop bulge whose structure had been determined by NMR. The resulting mRNA could be expressed and capped in a wheat germ lysate translation system supplemented with T7 polymerase following addition of m<sup>7</sup>G to the lysate (see, Figure 3A). Insertion of a 9-base leader before the TAR structure (HIVluc + 9) enhanced the translational 15 efficiency, presumably by allowing the pre-initiation complex to form. The helicase activity associated with the pre-initiation complex can transiently melt out the TAR RNA structure, and the message is translated (see, Figure 3A). Addition of a 39 amino acid *tat* peptide to the lysate stabilized the TAR RNA structure and inhibited the expression of the luciferase protein, as expected from a specific interaction between the TAR RNA and *tat* (see, Figure 3B).

20 "Small" organic molecules were then found that could inhibit the translation of the TAR-luciferase mRNA by stabilizing the TAR RNA structure. Compounds for the Available Chemicals Directory were docked to the TAR RNA structure and scored for binding energies. Among the best 25 compounds was ACD 00001199, whose structure is shown below. This compound has been shown to bind to TAR RNA with sufficient affinity to 25 disrupt the interaction with *tat* peptide at a 1  $\mu$ M concentration.

## ACD 00001199 Structure



Addition of 00001199 to the wheat germ lysate translation system with the luciferase mRNA produced some inhibition of translation at very high concentrations (see, Figure 4). However, the compound was much more efficient in inhibiting translation of the 5 luciferase mRNA containing the TAR RNA structure in the 5'-UTR, reducing translation by 50%, at a 50  $\mu$ M concentration. Small molecules that do not bind specifically to the TAR RNA structure did not affect translation of either mRNA construct.

**Example 6: Comparison of QXP predicted ligand-DNA structures to X-ray crystallography**

The utility of QXP in the context of ligands that bind to nucleic acid targets was evaluated. The X-ray data for netropsin (a minor groove binding drug) bound to two different duplex DNA sequences (PDB ID: 261d and 195d respectively (PDB IDs are identification codes for structures deposited in the Protein Data Bank, maintained at the Research Collaboratory for Structural Bioinformatics)) and an intercalator bound to an octamer duplex (PDB ID: 2d55) were used in validation studies. Root mean square (*rms*) deviations between the lowest energy docked structure (with randomly disordered ligands as initial structures) and the energy minimized X-ray structure fall within 0.6  $\text{\AA}$  in all the cases. Given that QXP method employs Monte Carlo type algorithm to search the conformational space and to make sure that the method is reliable in yielding global minimum, at least 10 QXP docking simulations were run with very different initial ligand structures. The performance of the QXP docking method can be quantified by its ability to identify the bound conformation of the ligand within 1.0  $\text{\AA}$  *rms* deviation from the crystallographically observed conformation. In the test cases described above, the success rate of the QXP runs is in the 80% range. The nearly linear correlation between the *rms* deviation from the crystal structure

and the score of the docked structure indicates that the QXP method is sufficiently accurate in predicting structures of ligand-receptor complexes.

**Example 7: Prediction of paromomycin-RNA complex structure using the QXP method**

The QXP method was used to derive an accurate structure of a bound ligand 5 to the RNA target. The NMR structure of the bacterial 16S ribosomal A site bound to paromomycin (Fourmy *et al.*, *Science*, 1996, 274, 1367; PDB ID: 1pbr) was used as the reference state. The aminoglycoside antibiotic was removed from the ligand-RNA complex. The conformation space of paromomycin was exhaustively searched using the QXP method 10 for the lowest energy conformers. The target RNA was held rigid whereas the paromomycin was treated as fully flexible. Multiple docking searches with the randomly disrupted paromomycin as initial structures were performed. The representative lowest energy structure identified from the search (dark grey) is superimposed on the NMR structure (light grey) of the bound complex as shown in Figure 5. The robustness of the QXP method is indicated (in Figure 6), through a correlation between the observed rms deviation and QXP energy scores.

**What is claimed is:**

1. A method of identifying compounds which bind to a molecular interaction site of a nucleic acid comprising:
  - providing a numerical representation of the three dimensional structure of said molecular interaction site;
  - providing a compound data set comprising numerical representations of the three dimensional structures of a plurality of organic compounds; and
  - comparing the numerical representation of the molecular interaction site with members of the compound data set to generate a hierarchy of said organic compounds, said hierarchy being ranked in accordance with the ability of said organic compounds to form physical interactions with said molecular interaction site.
2. The method of claim 1 wherein said ranked hierarchy identifies those compounds which bind to the molecular interaction site.
3. The method of claim 1 wherein the comparing is carried out *seriatim* upon the members of the compound data set.
4. The method of claim 1 further comprising chemically synthesizing said organic compounds which rank high in said hierarchy.
5. The method of claim 4 further comprising assessing the interaction of said highly ranked organic compounds with said nucleic acid.
6. The method of claim 5 wherein said assessment comprises mass spectrometry of a mixture of said nucleic acid and at least one of said compounds.
7. The method of claim 5 wherein said assessment comprises a functional bioassay.
8. The method of claim 1 further comprising generating a further data set of representations of organic compounds, said organic compounds comprising compounds which

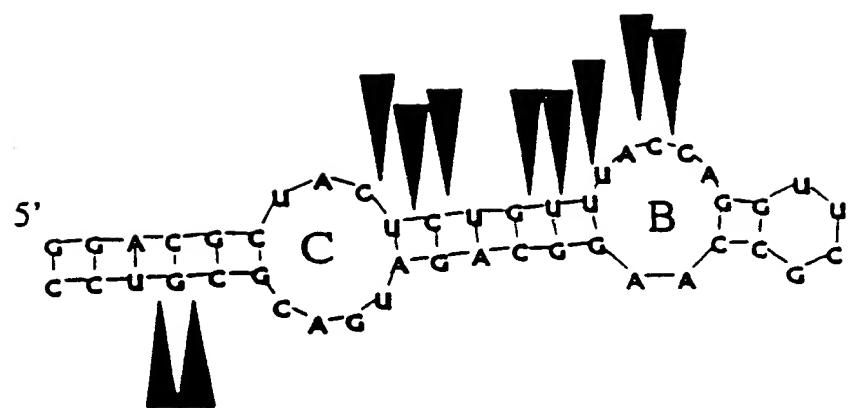
- 26 -

are chemically related to the organic compounds which rank high in said hierarchy, and comparing the numerical representation of the molecular interaction site with members of the further data set to determine a further hierarchy ranked in accordance with the ability of the organic compounds to form physical interactions with said molecular interaction site.

9. The method of claim 8 performed iteratively.
10. The method of claim 1 wherein said nucleic acid is RNA.
11. The method of claim 10 wherein said RNA is eukaryotic.
12. The method of claim 11 wherein said RNA is selected from the group consisting of mRNA, pre-mRNA, tRNA, rRNA, and snRNA.
13. The method of claim 10 wherein said nucleic acid is prokaryotic.
14. The method of claim 13 wherein said RNA is viral.
15. The method of claim 13 wherein said RNA is bacterial.
16. The method of claim 1 wherein said comparing is performed *in silico*.
17. The method of claim 1 wherein said molecular interaction site is present in a region of an RNA which is highly conserved among a plurality of taxonomic species.
18. The method of claim 1 performed for a plurality of molecular interaction sites.
19. The method of claim 1 wherein said molecular interaction site is located in the 3' or 5' untranslated region of a prokaryotic or eukaryotic mRNA.

- 27 -

20. The method of claim 1 wherein said molecular interaction site is located in the 5' untranslated region of mRNA associated with a disease process.
21. A data set comprising the numerical representations of the three dimensional structures of molecular interaction sites determined in accordance with claim 18.
22. A data set comprising the numerical representations of the three dimensional structure of organic compounds ranked high in the hierarchy generated in accordance with the method of claim 1.
23. A data set comprising the numerical representations of the three dimensional structures of a plurality of organic compounds in accordance with the method of claim 1.

**FIGURE 1**

ACD Code	Structure	Calc. $\Delta G$ of binding (kcal/mole)	$IC_{50}$ ( $\mu M$ )
000001199		-5.1	< 2
00192509		-8.5	< 2
000003934		-5.1	< 50

Figure 2

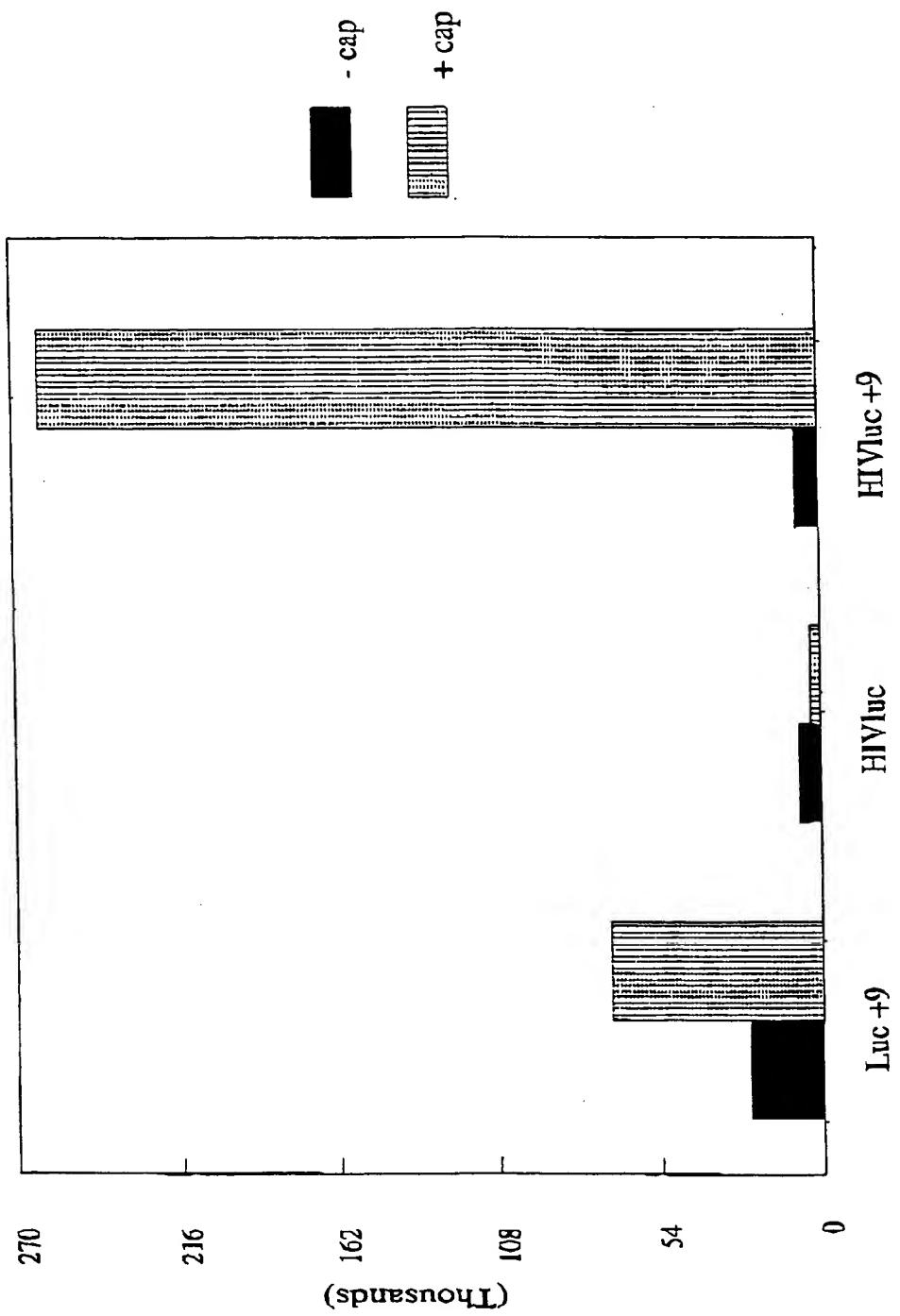
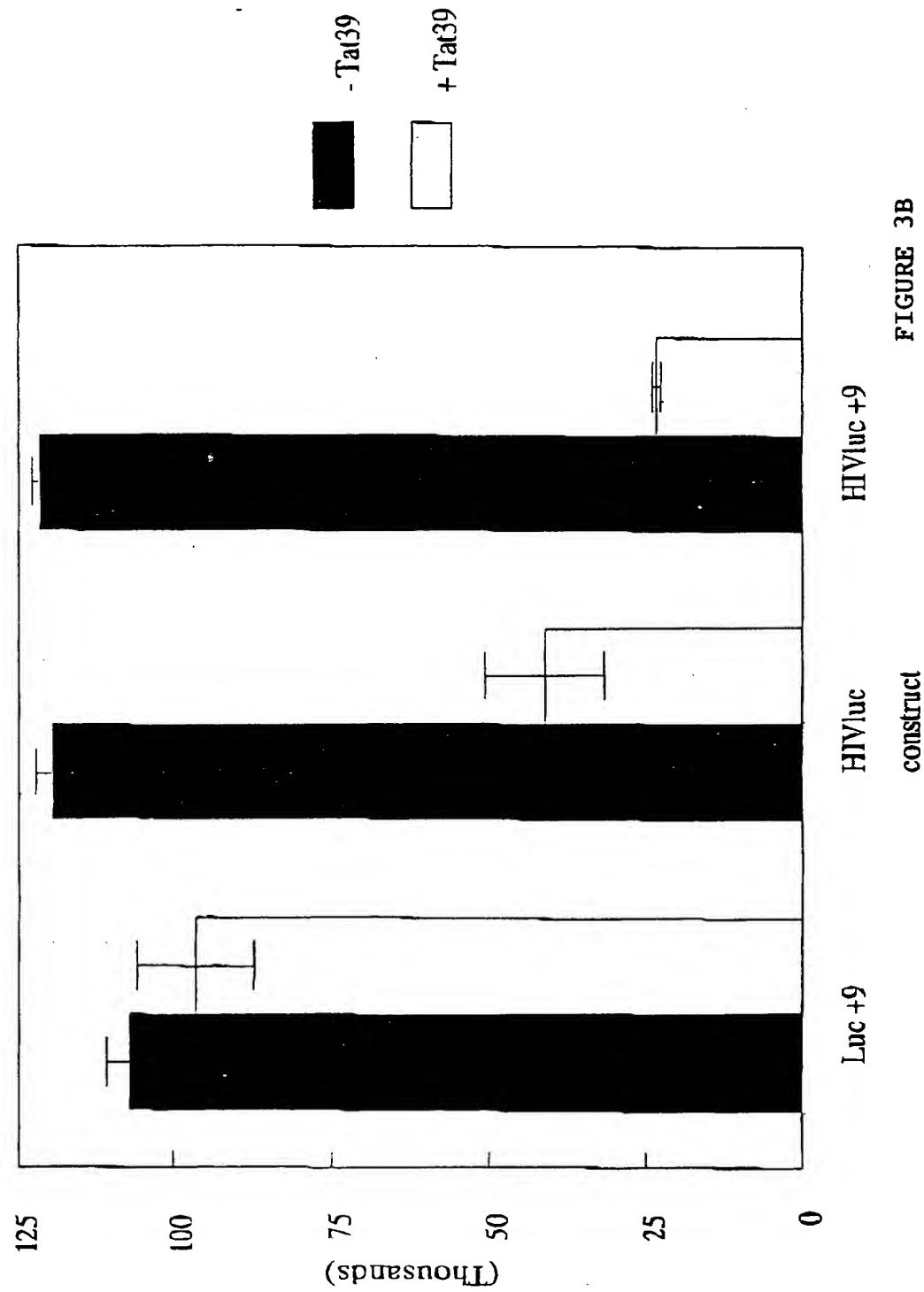


FIGURE 3A



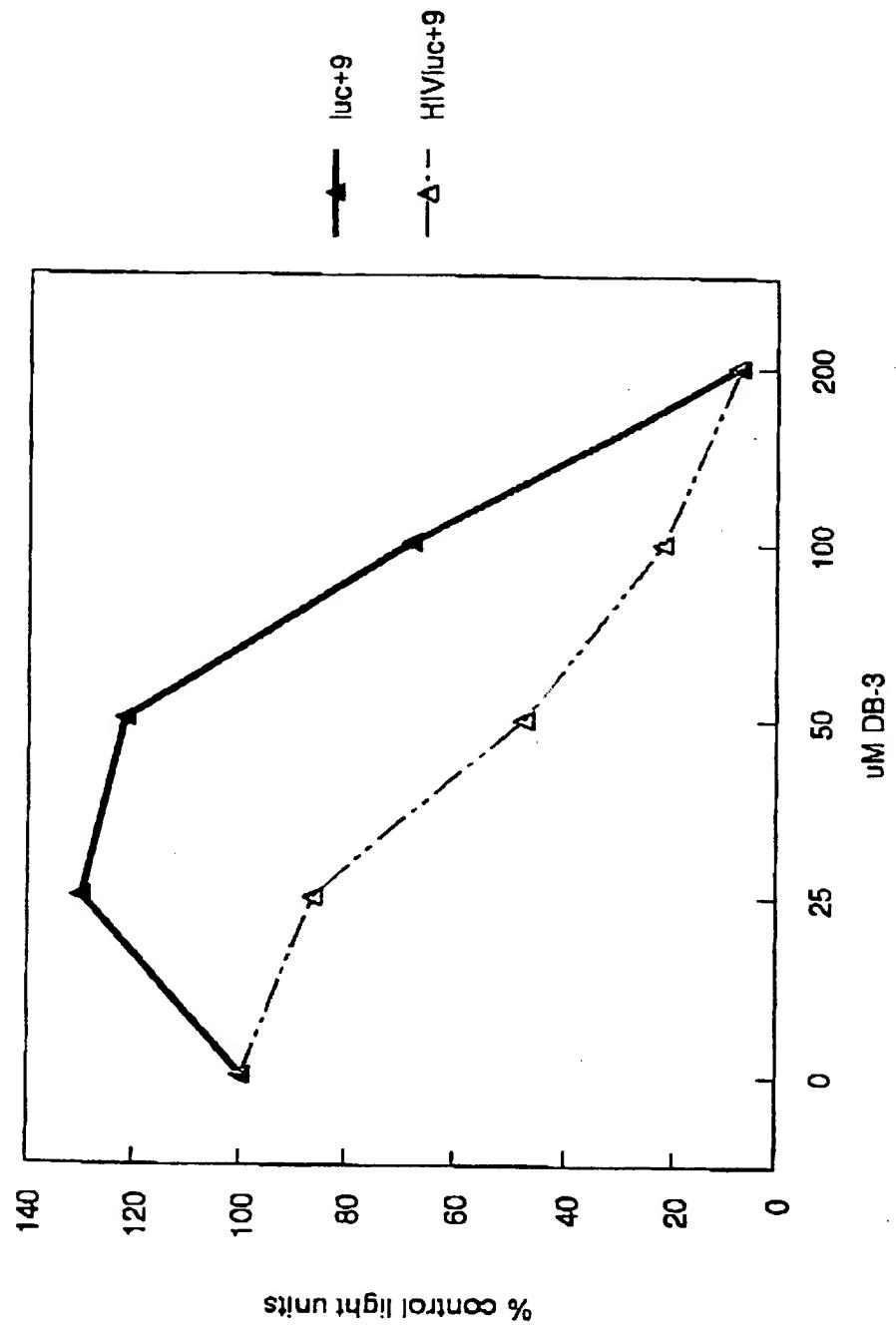
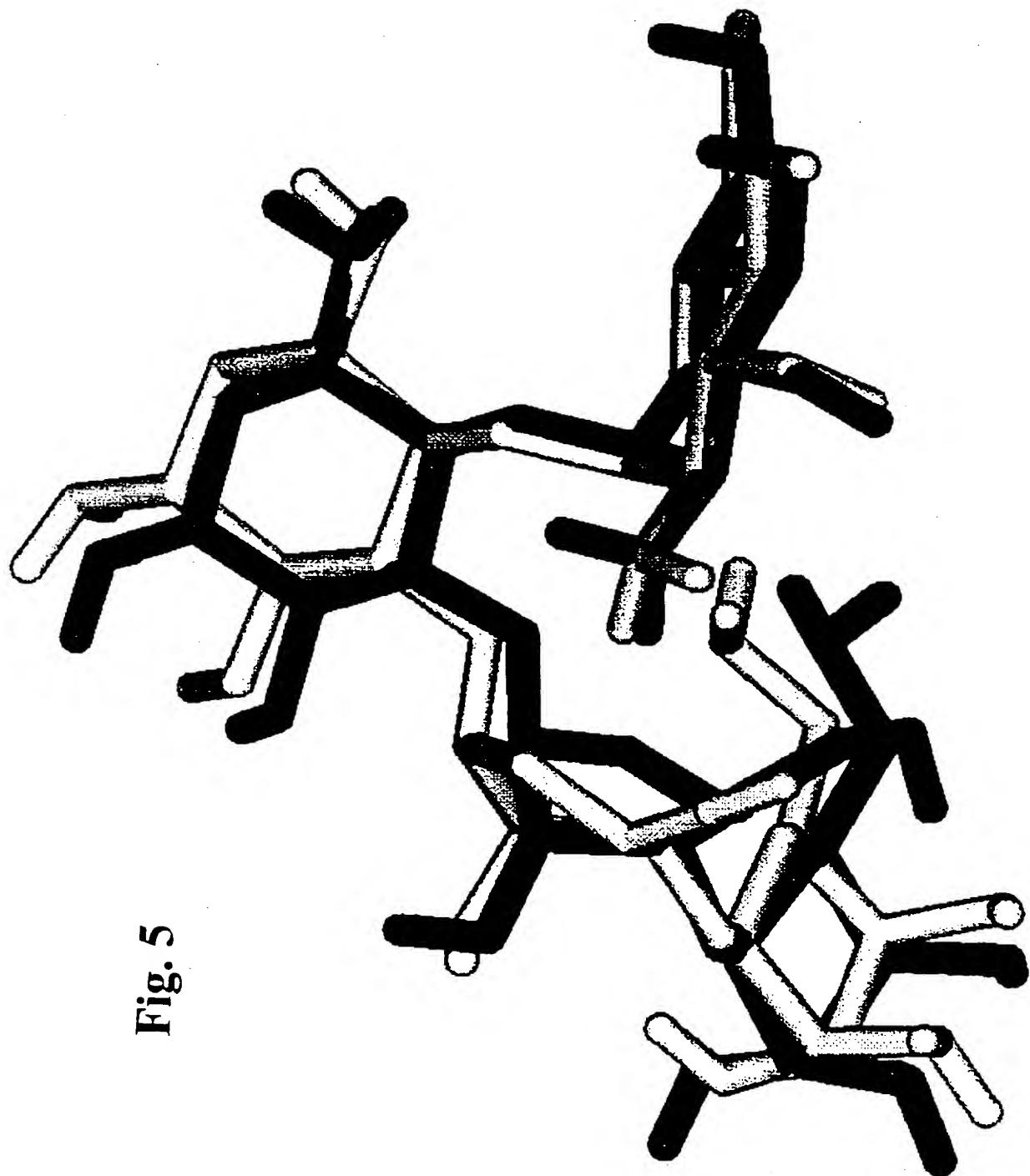


FIGURE 4



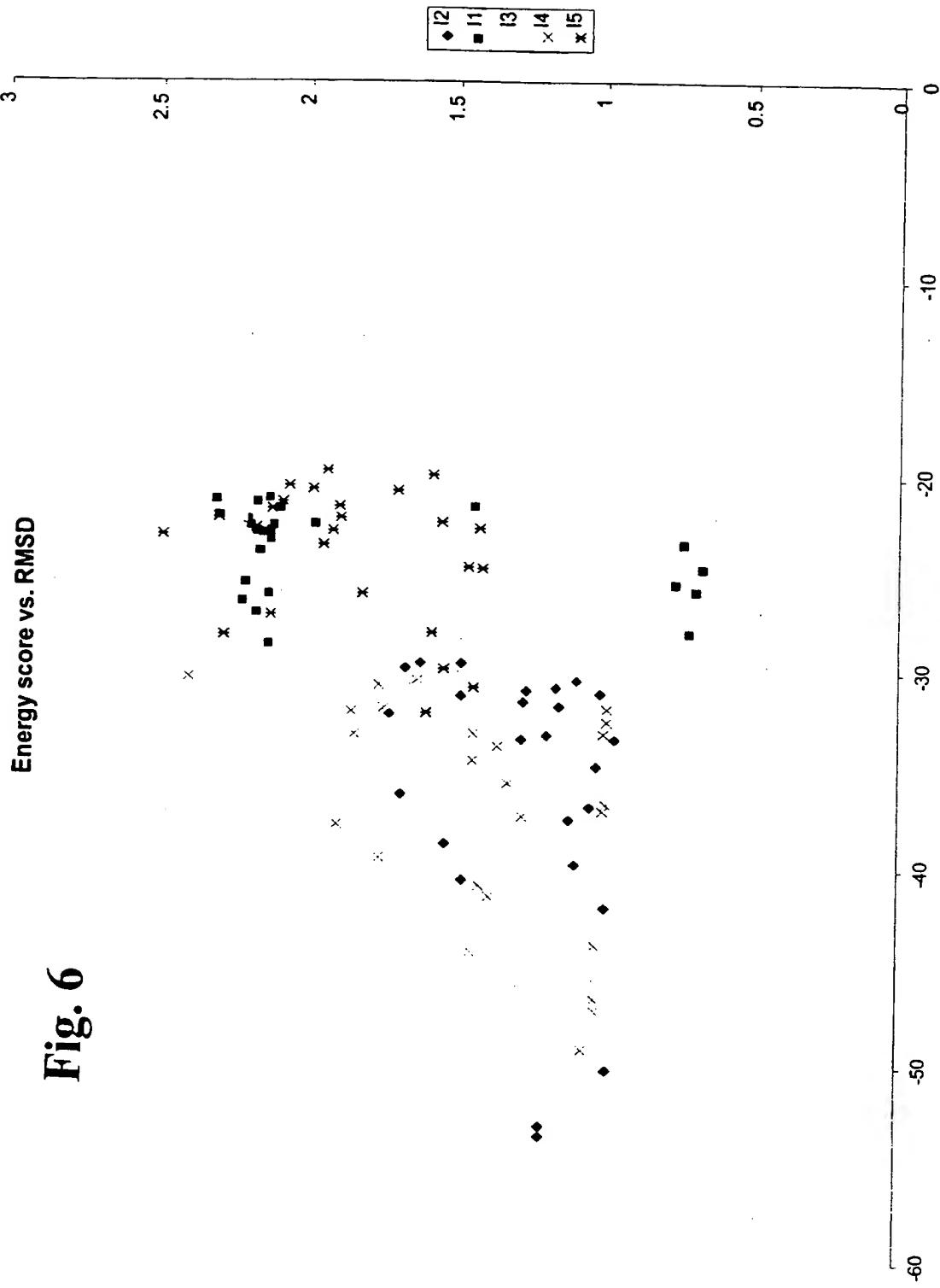


Fig. 6

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/10510

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; C12P 19/34; G05B 15/00; C07H 21/02, 21/04  
US CL : 435/6, 91.1; 702/19, 20; 536/22.1, 23.1, 24.3, 24.5

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1; 702/19, 20; 536/22.1, 23.1, 24.3, 24.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	US 5,888,738 A (HENDRY) 30 March 1999, see entire document.	1-5, 7-9, 16, 18
-----		-----
Y, P		6, 10-15, 17, 19-20
X	CHEN et al. Structure based discovery of ligands targeted to the RNA double helix. Biochemistry. September 1997, Vol. 36, No. 38, pages 11402-11407, see entire document.	1-5, 7, 10-16
Y		-----
Y	CRAIN et al. Applications of electrospray ionization mass spectrometry to structural studies of oligonucleotides and RNA. Abst. ACS. 1994, ANYL99, see entire document.	6, 8, 9, 17-20
		1-20

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance		
*E* earlier document published on or after the international filing date	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*O* document referring to an oral disclosure, use, exhibition or other means	*Z*	document member of the same patent family
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

07 JULY 1999

Date of mailing of the international search report

08 SEP 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No (703) 305-3230

Authorized officer

JEFFREY FREDMAN

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/10510

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BRADDOCK et al. Blocking of Tat-dependent HIV-1 RNA modification by an inhibitor of RNA polymerase II processivity. Nature. 04 April 1991, Vol. 350, pages 439-441, see page 439, column 1.	1-20
Y	RAMEZANI et al. Comparative analysis of five highly conserved target sites within the HIV-1 RNA for their susceptibility to hammerhead ribozyme mediated cleavage in vitro and in vivo. Antisense & Nucleic Acid Drug Development. 1996, Vol. 6, pages 229-235, see entire document.	1-20

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/10510

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: 21-23  
because they relate to subject matter not required to be searched by this Authority, namely:  

The claimed data sets are not proper subject matter since they represent mere presentation of information, which is not searched as per PCT Rule 39.
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

## Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/10510

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE, BIOSIS, CAPLUS

search terms: interact, bind, associate, hybridize, DNA, RNA, nucleic, polynucleotide, oligonucleotide, computer, program, numerical, representation, molecular, hierarchy, rank, list, mass, spectrometry